DOCUMENT RESUME

ED 385 578 TM 024 016

AUTHOR	Henning, Grant
TITLE	A Study of the Effects of Contextualization and Familiarization on Responses to the TOEFL Vocabulary Test Items.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-91-23; TOEFL-RR-35
PUB DATE	91
NOTE	71p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Adults; *Context Effect; Difficulty Level; *English (Second Language); Inferences; *Multiple Choice Tests; Reading Tests; Responses; Test Format; Test Items; Test Reliability; *Vocabulary Skills
IDENTIFIERS	Embedded Items; *Familiarization; *Test of English as a Foreign Language

ABSTRACT

In order to evaluate the Test of English as a Foreign Language (TOEFL) vocabulary item format and to determine the effectiveness of alternative vocabulary test items, this study investigated the functioning of eight different multiple-choice formats that differed with regard to: (1) length and inference-generating quality of the stem; (2) the nature of the task (matching versus supply); and (3) the degree of passage embeddedness of item stems or response options. In all, 1,040 vocabulary test items (80 familiarization and 960 experimental items) were developed and administered to 190 adult English-as-a-Second-Language students, 99 with and 91 without a prior familiarization activity. Results indicate that the current TOEFL vocabulary item format performed comparatively well in terms of difficulty, mean internal consistency reliabilty, and criterion-related validity. Among the alternative formats considered, only items embedded in reading passages appeared to outperform current TOEFL vocabulary format. Participation in the familiarization activity did not relate significantly or differentially to performance with any item type. Seven tables present analysis results. Three appendixes contain a sample test, sample familiarization materials, and the test taker questionnaire. (Contains 15 refrences.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 385 578

TOEFL

TEST OF ENGLISH AS A FOREIGN LANGUAGE

Research Reports

REPORT 35
FEBRUARY 1991

A Study of the Effects of Contextualization and Familiarization on Responses to the TOEFL Vocabulary Test Items

Grant Henning

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EDUCATIONAL TESTING SERVICE

TM024016

BEST COPY AVAILABLE

A Study of the Effects of Contextualization and Familiarization on Responses to the TOEFL Vocabulary Test Items

Grant Henning

Educational Testing Service
Princeton, New Jersey

RR-91-23

Educational Testing Service is an Equal Opportunity/Affirmative Action Employer.

Copyright © 1991 by Educational Testing Service. All rights reserved.

Unauthorized reproduction in whole or in part is prohibited
without written permission from the publisher.

TOEFL, the TOEFL logo, ETS, and the ETS logo are trademarks of Educational Testing Service,
registered in the U.S.A. and in many other countries.
Educational Testing Service is a U.S.-registered trademark.

Acknowledgments

Special acknowledgment is due the following persons for their substantial contribution to the success of this project:

Joanne Farr, for typing all items and tests for field testing and for designing results tables for the final report.

Thomas Ward, for assistance in writing the large number of test items employed in the study.

Marjorie Walsleben, for coordination and oversight of test administration at the UCLA Extension English Language Center.

Patricia Dunkel and Vicki Talbott, for arranging and coordinating test administration at Pennsylvania State University.

Nancy Parr, for editing the final report.

Mary Schedl, Karin Steinhaus, and Barbara Suomi, for advice and editorial insights on item writing procedures.

Peter Chilson, Patrice Dally, Gene Halleck, Elizabeth Henly, and Bruce Miller, for proctoring of experimental tests and familiarization activities.

The TOEFL Research Committee and TOEFL Committee of Examiners, for their encouragement and support of the project.

The Test of English as a Foreign Language (TOEFL) was developed in 1963 by a National Council on the Testing of English as a Foreign Language, which was formed through the cooperative effort of more than thirty organizations, public and private, that were concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program, and in 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the Graduate Record Examinations (GRE) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education.

ETS administers the TOEFL program under the general direction of a Policy Council that was established by, and is affiliated with, the sponsoring organizations. Members of the Policy Council represent the College Board and the GRE Board and such institutions and agencies as graduate schools of business, junior and community colleges, nonprofit educational exchange agencies, and agencies of the United States government.

A continuing program of research related to the TOEFL test is carried out under the direction of the TOEFL Research Committee. Its six members include representatives of the Policy Council, the TOEFL Committee of Examiners, and distinguished English as a second language specialists from the academic community. Currently the Committee meets twice yearly to review and approve proposals for test-related research and to set guidelines for the entire scope of the TOEFL research program. Members of the Research Committee serve three-year terms at the invitation of the Policy Council; the chair of the committee serves on the Policy Council.

Because the studies are specific to the test and the testing program, most of the actual research is conducted by ETS staff rather than by outside researchers. However, many projects require the cooperation of other institutions, particularly those with programs in the teaching of English as a foreign or second language. Representatives of such programs who are interested in participating in or conducting TOEFL-related research are invited to contact the TOEFL program office. Local research may sometimes require access to TOEFL data. In such cases, the program may provide the data following approval by the Research Committee. All TOEFL research projects must undergo appropriate ETS review to ascertain that the confidentiality of data will be protected.

Current (1990-91) members of the TOEFL Research Committee are:

Patricia L. Carrell (Chair)	University of Akron
James Dean Brown	University of Hawaii
Patricia Dunkel	Pennsylvania State University
Fred Genesee	McGill University
Elliott Judd	University of Illinois at Chicago
Elizabeth C. Traugott	Stanford University

Abstract

In an effort to evaluate criticisms of the current TOEFL® vocabulary item format and to ascertain the comparative effectiveness of alternative vocabulary test item formats, the present study investigated the functioning of eight different multiple-choice formats that differed with regard to the length and inference-generating quality of the stem, with regard to the nature of the task (matching versus supply), and with regard to the degree of passage embeddedness of item stems or response options. In all, 1,040 vocabulary test items (80 familiarization items and 960 experimental items) were developed and administered in counterbalanced clusters to a total participating sample of 190 adult English-as-a-second-language students (99 persons with a controlled prior 75-minute familiarization activity and 91 with no familiarization activity) at two language institutes. These two levels of familiarization were examined to ascertain effects of familiarization on performance with the eight item types. Also, self-reports of levels of prior exposure to particular item types were correlated with actual performance with those item types.

Results indicated that, for the present sample of persons and items, the current TOEFL vocabulary item format performed comparatively well in terms of appropriateness of difficulty, mean internal consistency reliability, and criterion-related validity. Such performance came in spite of earlier suggestions that current item stems may be too long, lack inference-generating information about the meaning of the targeted vocabulary, and place undue reliance on matching rather than supply format.

Among alternative item formats considered, only items embedded in reading passages appeared to outperform current TOEFL vocabulary format in terms of estimates of both reliability and criterion-related validity. Use of an item format that differed from current TOEFL vocabulary format only in the uniform presence of inference-generating information in the item stem and in reduction of mean length of stem also slightly outperformed current TOEFL vocabulary item format in terms of the particular criterion validity indicated by correlation with vocabulary total score; however, these correlational differences did not quite attain statistical significance. In general, multiple-choice items using stems with underscored target vocabulary in context performed better than did either multiple-choice single-word-or-phrase matching tasks or multiple-choice supply-type items.

Participation in the familiarization activity did not relate significantly or differentially to performance with any item type. However, self-reports of familiarity with particular item types suggested that some types were more familiar than others, and the three most familiar item types, including current TOEFL vocabulary format, exhibited a significant positive correlation between self-report of familiarity with the item type and successful performance with the item type.

PROBLEM

It has long been considered superior language testing procedure to present vocabulary test items in an authentic context rather than to present them as isolated words for mere matching tasks (Oller, 1979; Stansfield, 1986). The presumed advantage of contextualization is in the likelihood that it enhances face validity, inasmuch as contextualized tasks more nearly approximate natural language use. Also, for lexical items with multiple definitions, context may serve to delimit the intended meaning for testing purposes. To accommodate the need for contextualization, the current TOEFL vocabulary item format of the vocabulary section presents lexical items underscored in extensive syntactic context. (See example item 1 below.) The obvious logic of providing such context is at least threefold: (1) it answers the criticism of need for face validity since a more natural environment for encountering lexical items is thereby presented; (2) it occasionally allows for the possibility of actual inference of meaning from the verbal environment; and (3) it causes reading and vocabulary formats to become more compatible and more logically grouped within the same section since extensive reading is invited for both the vocabulary and reading formats.

Pike (1979), in a systematic study of prospective TOEFL item types, found that multiple-choice measures of vocabulary were among the most efficient item types investigated; however, his study considered only synonym matching and sentence completion type items among the vocabulary item formats compared. Those item types corresponded to two of the eight item types considered in the present study--that is, isolated word or phrase multiple-choice matching format, and minimal sentence stem multiple-choice cloze-supply format. (See sample items 2 and 4 below.) Unfortunately, those two item types were the two types entailing the least contextualization by comparison with any of the item types considered in the present study.

In a study of TOEFL dimensionality across levels of examinee proficiency DeJong and Henning (1990) found the current TOEFL vocabulary section to be among the most consistent of the TOEFL item types in terms of its strength of contribution to the measurement of proficiency across four controlled levels of language proficiency.

Despite the utility of the current TOEFL vocabulary item format, use of that format has not been without criticism (Duran, Canale, Penfield, Stansfield, and Liskin-Gasparro, 1985; Stansfield, 1986). One functional problem associated with the current TOEFL contextualized vocabulary format is that the item sentence stems are so long that a mere reversion to matching of the underscored word with a provided option may become a more efficient responding strategy than taking time to read the entire sentence. This is particularly true in the case of the many items that have extensive context in the stem but for which the context sheds little or no light on the meaning of the vocabulary tested. (See example item 1 below.) Thus, it may well be that the use of overly extensive context

provides an unfair advantage to testwise students for this section of the test, and responding to the items may not uniformly require the same intended skills of all examinees. Examinees who expend the time necessary to read each sentence may actually obtain lower scores than they would have had they used a simple matching strategy because, as a result of their approach, they may not have sufficient time to finish all items. This problem might be avoided through the use of possible improvements in item format, such as reduced context in the item stem, or cloze supply techniques rather than synonym matching. Thus, a focus of this study has been to determine whether the current TOEFL vocabulary section can be improved through use of a format that alters the item contextualization in degree or the response task in kind.

Another potential problem related to the contextualization problem described above pertains to the inference-generating quality of the context provided in the sentence stem. With the current vocabulary item format, some stems provide sufficient information for possible inference of the meaning of the targeted vocabulary word or phrase, and some do not provide enough information. (See sample items 1 and 5 below.) Several language testing specialists have considered the ability to arrive at lexical meaning through context cues as an important skill deserving measurement focus (Oller, 1979; Valette, 1977). Williams and Dallas (1984) reported that inability to draw correct meaning inferences when sufficient contextual information is present is especially characteristic of less proficient second language learners. It is worth empirical study to ascertain whether items that permit appropriate inference generation are superior to items that do not permit such inference of meaning from context. To this end, in the present study, two of the formats compared were specifically designed to present inference-generating information in the item stem that could help inference of the meaning of the targeted vocabulary. A third format was designed to embed the sentence stems in reading passages so that inferencing information might be available both within and beyond the sentence in which the targeted lexis appeared.

A further potential problem with existing or alternative item formats is related to the degree of familiarity of such formats and, more particularly, the impact of familiarization on performance. TOEFL examinees come from a vastly differing array of linguistic, cultural, and educational backgrounds (Wilson, 1982). Many come from backgrounds that make it unlikely that they would have had prior exposure to certain test item formats. If their lack of exposure to a type of vocabulary item format causes them to perform less successfully than other examinees who have had such exposure independent of their comparative knowledge of English vocabulary, then lack of exposure becomes a disadvantage associated with test format rather than with knowledge or lack of knowledge of vocabulary. To some extent, and for some examinees, these differential exposure problems are overcome through the distribution of sample items in the TOEFL test information literature. It would, nevertheless, be useful in investigating comparative item format quality to determine whether performance with particular formats is differentially affected by the opportunity for familiarization provided. Presumably, formats that are least susceptible to familiarization effects

would, other things being equal, be most suitable for the testing of EFL vocabulary.

FOCUS

The present study investigated the comparative functioning of eight vocabulary item types under two conditions of task familiarization. Criteria for successful item functioning included comparative internal consistencies of subtests comprised of each item type, validity of item types as indicated by subtest correlations with vocabulary total scores across all item types, nonsusceptibility to familiarization effects, and comparative mean difficulties and discriminabilities of items considered by item type. The eight vocabulary item format types investigated were the following:

1. The current TOEFL extended context vocabulary test item format as illustrated in the example below

Example (actual TOEFL item):

Human facial expressions differ from those of animals in the degree to which they can be <u>deliberately</u> controlled and modified.

(A) both
(B) noticeably
(C) intentionally
(D) absolutely

2. An isolated word or phrase, multiple-choice matching format

Example (adapted from actual TOEFL item in 1 above):

<u>deliberately</u>
(A) both
(B) noticeably
(C) intentionally
(D) absolutely

3. A minimal sentence stem (sufficient only to predict part of speech), multiple-choice matching format

Example (adapted from actual TOEFL item in 1 above):

He spied on them <u>deliberately</u>.
(A) both
(B) noticeably
(C) intentionally
(D) absolutely

4. A minimal sentence stem (as in 3 above), multiple-choice cloze-supply format

Example (adapted from actual TOEFL item in 1 above):

He planned the crimes _____.
(A) both
(B) noticeably
(C) intentionally
(D) absolutely

5. A reduced-length inference-generating sentence stem, multiple-choice matching format

Example (adapted from actual TOEFL item in 1 above):

He was guilty because he did those things deliberately.
(A) both
(B) noticeably
(C) intentionally
(D) absolutely

6. A reduced-length, inference generating sentence stem, as in 5 above, except with multiple-choice cloze-supply format

Example (adapted from actual TOEFL item in 1 above):

He was guilty because he did those things _____.
(A) both
(B) noticeably
(C) intentionally
(D) absolutely

7. A single word or phrase stimulus with passage-embedded multiple-choice options

Example (adapted from actual TOEFL item in 1 above):

Deliberately
(A) (B) (C)
They both enjoyed the show noticeably, and he intentionally

(D)
indicated that it was absolutely the best performance he had seen.

8. A sentence stem embedded within an extended reading passage, with multiple-choice matching format

Example (adapted from actual TOEFL item in 1 above):

In a (1) democratic society suspected persons are presumed innocent until proven guilty. The (2) establishment of guilt is often a difficult task. One consideration is whether or not there remains a (3) reasonable doubt that the suspected persons committed the acts in question. Another consideration is whether or not the acts were committed (4) deliberately. Still another concern is whether or not the acts were (5) premeditated.

4. (A) both
   (B) noticeably
   (C) intentionally
   (D) absolutely

The two levels of test familiarization considered were as follows:

1. No prior exposure to or familiarization with the particular test formats employed in this study was provided for approximately half of the subjects.

2. Prior exposure to and practice with sample tests and answer sheets using the identical item formats (although not using the same vocabulary) as the experimental tests was provided for the other half of the subjects. A practice test consisting of 80 items, 10 of each format type, was administered within a one-hour period on the day before experimental testing. Feedback was provided that included a list of correct options for each item.

METHOD

Sample

One hundred and ninety ESL students enrolled in intensive English language instruction programs at two major U.S. universities (UCLA Extension and Pennsylvania State University) participated in the study by

informed consent. Sites were chosen so as to represent a broad range of English language proficiency, age, and native language background as might be typical of the usual TOEFL examinee population. Actual demographic information about the sample, including language background, sex, age, and time of residence in the United States, is reported in Table 1.

Because participation was considered relevant to their program of instruction, subjects took part in the study during regular class time. In addition to practice in responding to a variety of vocabulary test types, participants were provided with either specially prepared vocabulary building materials or TOEFL practice kits as compensation and further incentive for participation.

Instrumentation

Eight experimental tests were developed, each one consisting of 120 items employing the same targeted vocabulary and multiple-choice options from past disclosed TOEFL forms, but each test presented the targeted vocabulary and response options in different item formats and in a different controlled and counterbalanced sequence (See Table 2 for an indication of the counterbalanced modular design of the experimental tests, and see Appendix A for a sample experimental test.) Thus, although 960 total items were developed and assembled, no person responded to more than one experimental test consisting of 120 items. Each test was composed of eight 15-item components corresponding to the eight item format types described in the purpose section above. The first group of 15 items for each experimental test consisted of actual disclosed TOEFL vocabulary test items. The remaining 105 items in each test were developed, in each case, by revising disclosed TOEFL vocabulary test items to conform to the seven alternative formats described above.

Because of the need to preserve the same multiple-choice response options for any given lexical item across the eight item formats, item writing for this experimental study was more challenging than would be the case in an operational item-writing context. For this reason, comparisons of the time required to write experimental items of the formats proposed would be uninformative of actual time required in an operational context. For this reason also, the language of some of the items in the appendix is somewhat less authentic and more stilted in appearance than would be the case in an operational context. The same severe constraint also resulted in a few items having multiply-keyable options in spite of careful editing by several individuals. Where it was recognized that this problem could occur, instructions were provided to choose the best or most common option. It was believed that there was a best choice in every case and that persons with the knowledge tested would be able to find that best choice. The fact that one particular item type (i.e., minimal sentence stem, multiple-choice supply) was more susceptible than others to this problem was reflective of a weakness inherent to that item type that would be difficult to overcome even in an operational context. Therefore, it did not seem inappropriate that the

differential tendency for this weakness to occur should be allowed to persist in the design of the study.

An 80-item sample test was developed by adaptation of previously disclosed TOEFL items, consisting of 10 items from within each of the eight format types described above (See Appendix B for a copy of the familiarization test.) This 80-item test battery and its accompanying answer sheet were used for familiarization purposes for those 99 subjects randomly assigned to the familiarization treatment. None of the targeted lexical items for this familiarization test were the same as those employed in the experimental tests described in the preceding paragraph.

A demographic questionnaire was used with all subjects to elicit native language, sex, age, length of U.S. residence, and self-report of degree of familiarity with item formats appearing on the experimental test (See Appendix C for a copy of the questionnaire.) This self-report of familiarity provided further information about the effects of prior exposure to the item types, in addition to the information made available through use of the familiarization task. Familiarity items requested self-report of degree of familiarity with the eight item formats on a four-point scale.

Procedures

The 190 subjects were assigned by random stratification on proficiency level, as assessed by prior placement testing, to the two levels of familiarization--99 subjects to the familiarization condition and 91 subjects to the nonfamiliarization condition. Table 2 presents a summary of the design and procedures of the study taking place on two successive days in each site.

Familiarization. On the first day, for those 99 subjects selected for the familiarization activity, 75 minutes were devoted to format familiarization. Ten minutes were used for instructions, 60 minutes for responding to the 80-item practice test, and the final 5 minutes were used to provide feedback on performance. Feedback consisted of dissemination of written answers to all vocabulary items and responding to oral questions. Subjects were allowed to keep the answer sheets, but the practice test booklets were not returned to them until the end of the second day in order to ensure that none of the subjects in the nonfamiliarization condition would have the benefit of the same exposure to the experimental item formats.

Experimental Testing. On the second day, a total of 179 persons appeared for experimental testing at the two sites. A total of 115 minutes were allocated for experimental testing as follows: 10 minutes for instructions, 5 minutes for the demographic questionnaire, 10 minutes for the familiarity questionnaire, and 90 minutes for the 120-item experimental vocabulary test forms. Test booklets were distributed by a systematic spiraling procedure that ensured approximately equal numbers of persons responded to each one of the eight test forms. The manner in

which test booklets were disseminated also ensured randomized assignment of test booklets to subjects. On average, each test booklet was administered to a randomized subgroup of 22.38 persons. Groups ranged in size from 19 to 24 persons, as indicated in Table 2. Each group was administered one test booklet consisting of 120 items in eight format groups of 15 items each. While each group received the same numbers of test items over the same eight format types and over the same 120 lexical items and their associated options, actual vocabulary words tested were systematically counterbalanced across the eight format types. In this way no subject encountered the same targeted vocabulary word twice, but every targeted vocabulary word was employed in every experimental format to test approximately the same number of persons as every other targeted vocabulary word. Targeted vocabulary items were grouped in 15-item clusters under each item type, and these clusters were presented in counterbalanced sequence such that every vocabulary cluster appeared in the same sequence the same number of times as every other vocabulary cluster. To further control for any possible sequence effect, item groups were numbered sequentially in the order in which they were encountered. Correlations of encounter sequence with performance on item groups approached zero and were nonsignificant within and across all person groups.

All subjects were provided ample time to respond to every item within the 90 minutes allowed for 120 items. Subjects were allowed to leave the testing room on completion of the test. Even the last remaining subject was permitted time to respond to every item. In this way, the possibility of any speededness effect was either eliminated altogether, or was greatly reduced.

Analyses

a. Means, standard deviations, intercorrelations, and internal consistency reliabilities were computed for all tests and item format subtest groupings for the 179 subjects who responded to all experimental vocabulary tests and for the 172 subjects who responded to the familiarity questionnaire.

b. Item analyses including estimation of item difficulty (p and Delta) and item discrimination (Fisher-z-transformed item-total point biserial correlation) were conducted for all experimental vocabulary items in each test format.

c. One-way ANOVA was used to test the significance of format effects on item difficulty estimates and on item discriminability estimates for the 960 experimental vocabulary items. Post hoc analyses were used to test differences between pairs of difficulty means and between discriminability means for each item format type. Analysis of covariance was used to analyze those same

format effects after control was made for self-reported familiarity levels with each item type. A test of ANCOVA within-cell regression assumptions indicated that ANCOVA was appropriate as a test of discriminability mean differences, but not as a test of difficulty mean differences.

RESULTS

Subtest Difficulty Comparisons

Table 3 reports mean scores across the eight person groups and eight 15-item clusters for every item format type. Means ranged from a low of 6.832 for the most difficult item format (i.e., reduced-length inference-generating sentence stem, multiple-choice cloze-supply format) to a high of 8.821 for the easiest item format (i.e., minimal sentence stem, multiple-choice matching format). However, all of the item format types appeared to fall within an acceptable range for difficulty. The current TOEFL vocabulary item format ranked second for easiness, with a mean of 8.335. Judging from mean comparisons, there was a predictable tendency for matching formats to be easier than supply formats.

Subtest Reliability Comparisons

Conservative estimates of internal consistency reliabilities within format types were provided by use of the Kuder-Richardson Formula 21 procedure. Mean estimates across eight different 15-item vocabulary clusters in each format type were adjusted by use of the Spearman Brown Prophecy formula to provide conservative estimates of internal consistency reliability for a 120-item test in each format type. As Table 3 indicates, use of a passage-embedded sentence stem with multiple-choice matching format made for the most reliable assessment among the eight alternative item types investigated. Current TOEFL vocabulary item format ranked second among formats considered, with regard to associated estimates of internal consistency reliability. There was a tendency for matching formats to be associated with higher estimates of reliability than supply formats. And there was a similar tendency for inference-generating formats to be associated with higher estimates of reliability than formats in which the item stems did not supply information to assist in the inference of the meaning of the targeted vocabulary. Reference to the normal distribution of Fisher z suggested that, for the present sample, a coefficient of 0.947 could be said to exceed a coefficient of 0.926 at a one-tailed alpha level of 0.05. Since the estimated reliability for the current TOEFL format was 0.927, it would appear that the estimated difference in reliability for the two formats was very nearly above the level of chance.

Subtest Validity Comparisons

Table 5 reports the intercorrelations among item-cluster totals for all item format types, with total scores across all item types and with two measures of familiarity with item types--participation or

nonparticipation in the familiarization activity (FA), and self-reported level of familiarity with a particular item type (Q). The highest correlation with total vocabulary score was obtained by the passage-embedded sentence stem, multiple-choice matching format (0.820). The second most valid format by this criterion was the reduced-length inference-generating stem, multiple-choice matching format (0.806). The current TOEFL vocabulary item format ranked third among candidate item formats in terms of validity by this criterion (0.793). Again, there was a tendency for comparable matching formats to surpass supply formats with regard to validity by this criterion (0.782 and 0.806 versus 0.708 and 0.658). It should be acknowledged that the criterion itself was composed of more matching item formats than supply formats, which could have provided an inherent advantage for matching formats; however, it can be observed that the individual matching formats correlated even more highly with the supply formats than the supply formats correlated among themselves, further supporting a generalization of validity superiority on the part of matching formats. Inference-generating matching format (ET) appeared to surpass non-inference-generating matching formats (AT and CT), suggesting a probable validity superiority for certain vocabulary item formats that provide inferencing information in the stem. It should be noted, however, that reference to Hotelling's t test for correlated coefficients of correlation indicated that, for a one-tailed test of the significance of the difference in these correlations at an alpha level of 0.05, only those coefficients of the magnitude of 0.765 and below could be said to differ significantly from the coefficient of 0.820 observed for the format involving passage-embedded sentence stems.

Familiarity Relationships

As indicated by the near-zero correlations in Table 5 for the binary-coded familiarization activity (FA) with cluster totals for individual item types (AT through HT), the 75-minute familiarization activity provided did not appear to affect performance differentially on vocabulary items of any of the formats considered here. While that is probably good news for test developers, it leaves unanswered the questions of what level of familiarization might indeed affect performance, and whether such effects would differ across item types. Since correlations ranged from -0.109 to 0.015, it is likely that the familiarization activity as conducted was not sufficient in duration, intensity, or focus to affect performance in responding to any of the present item types. Still, it is unlikely that highly extensive familiarization activities would be more suitable than the present procedure because subsequent changes in performance under conditions of extensive training might become confounded with differential learning of English rather than only reflecting increased familiarity with the methods of testing.

Self-reports of familiarity with item types (variable row Q in Table 5, also reported in column R of Table 4) were more informative of possible existing relationships with performance on item types than was the impact of the familiarization activity. Performance on current TOEFL vocabulary item format was positively related to self-reported familiarity with that item type (0.242, $p<0.01$, one-tailed test). Also,

performance on formats CT and ET (i.e., minimal sentence stem multiple-choice matching format and reduced-length inference-generating stem multiple-choice matching format) was positively related to self-reported familiarity with those formats (0.182 and 0.219, respectively, $p<0.05$, one-tailed test). It is also apparent from the means in Table 4 that those three item types, for which self-reported familiarity was positively related to performance with the item type, were also the three item types most familiar to these subjects. The observed relationship between self-reported familiarity and performance, then, may have been a function of the length and quality of the prior educational experience. Those who knew English best had more exposure to vocabulary tests, and the kinds of tests most frequent in their experience employed the three item types just mentioned.

A further analysis of the relationship between self-reported familiarity with item format type and performance with item type was provided by use of a Spearman rank correlation between facility means and familiarity means for the eight format types. A significant correlation (rho = 0.833, $p<0.01$) indicated a strong positive relationship between self-report of familiarity with an item format type and actual success in responding to that item type. However, it should be noted that the differences between means used to establish ranks were small and nonsignificant in many instances.

Following appropriate tests of the homogeneity of regressions, analysis of covariance was employed to test the significance of differences among means of item difficulty and differences among means of item discriminability for the 960 experimental items and eight format types, using self-reports of familiarity with item format type as a concomitant variable. In the case of item difficulty estimates, the assumption of homogeneity of regressions was not met, so the ANCOVA results are not reported. In the case of item discriminability estimates, the assumptions were met, but the impact of self-reported familiarity on item discriminability was so small that results of ANOVA reported in Table 6B were virtually unchanged.

Item Difficulty Comparisons

Table 6A reports the results of ANOVA to examine mean differences in item difficulties computed for the eight item formats and 960 items investigated. Item difficulties were computed as Delta statistics, where Delta = 13 - 4z, and z represented the standardized proportion correct value (p) for each item. The purpose of the Delta transformation was to normalize or standardize the proportion correct values to form an equal-interval scale, and to express difficulty as a positive value, unlike the situation with the scale of p values.

Results indicated that a highly significant difference existed among the means of item difficulty considered for the eight item formats investigated ($p<0.001$). At the same time, the multiple correlation of 0.192 suggested that a comparatively small portion of the overall variance in item difficulty was attributable exclusively to item format

type. Post hoc analyses are presented in Table 7. There it is evident that, not surprisingly, mean item difficulties were ranked in exact conformity to the difficulty rankings for the eight subtests reported in Table 3. Once again, the current TOEFL vocabulary item format was indicated to be the third easiest of the formats considered. Minimal sentence stem, multiple-choice matching format was the easiest format overall, and reduced-length inference-generating sentence stem, multiple-choice supply format was indicated as the most difficult item format. Tukey's HSD test of the critical range for pairs of Delta means (Einot & Gabriel, 1975) suggested that differences between individual pairs of means would need to exceed the critical value of 1.539 in order to be interpreted as differing beyond the level of chance ($p<0.05$). By this criterion, matching formats 3 and 5 could be said to be significantly easier than comparable supply formats 4 and 6, when considered in tandem.

Item Discriminability Comparisons

The ANOVA reported in Table 6B indicated that a highly significant difference in means of item discrimination indices existed when considered across the eight item formats ($p<0.001$). Again, however, it should be noted that the associated multiple correlation of 0.176 suggested that only a comparatively small portion of the variance in item discrimination was attributable to item format type. It should be noted here that the computations employed Fisher z transformations of item-total point biserial correlations. This transformation was used to convert existing correlations from an ordinal to an equal-interval scale for computation purposes. Correction for part-whole overlap was not done since the contribution of individual items to the 120-item test total was believed to be miniscule. Point biserial instead of the usual biserial item-total correlations were employed since, by design, every person group responded to every item type, so the comparability of point-biserials was not jeopardized by differences in abilities of responding subject groups (Gulliksen, 1987).

Post hoc analyses presented in Table 7 indicated that the rankings of item format type for comparative discriminability were very similar to the rankings for subtest reliability presented in Table 3. Again, the most discriminating format (0.414) was that of sentence stems embedded in reading passages with multiple-choice matching format. The least discriminating format (0.271) was that of minimal sentence stem, multiple-choice supply format. Current TOEFL vocabulary format ranked second in discriminability among the eight formats considered (0.378). Use of simple word or phrase matching format tended not to be so discriminating as matching format that employed whole sentence stems (0.298 versus 0.378, 0.367, 0.372, and 0.414). Tukey's HSD test of the critical range for pairs of z-transformed point biserial correlations indicated that a critical value of 0.106 would need to be exceeded in order for the difference between means for any pair to be significant above the level of chance ($p<0.05$).

DISCUSSION

The present study was conducted to compare the functioning of English vocabulary test items constructed according to eight explicit item format types, including the current TOEFL vocabulary item format and seven alternative formats. Alternative formats were chosen so as to permit comparison of differing levels and types of contextualization of targeted vocabulary, differing levels of inferencing information provided in the item stems, matching versus supply type multiple-choice response tasks, and use versus nonuse of embedding of item sentence stems in a reading passage. In all, 960 experimental items were prepared and administered in a controlled and counterbalanced manner to a sample of 179 English-as-a-second-language students enrolled in intensive English language institutes at two American universities.

In addition to traditional measures of test and item functioning, the study was designed to provide information about the comparative impact of familiarity with the respective item formats. Familiarity was assessed both by means of participation or nonparticipation in a 75-minute familiarization activity and by means of questionnaire self-reports of familiarity with the eight item formats considered.

Results indicated that the current TOEFL vocabulary item format functioned comparatively well among the alternative formats in terms of appropriateness of difficulty, estimated internal consistency reliability, criterion-related validity, and observed item discriminability. However, use of sentences embedded in reading passages tended to surpass the current TOEFL vocabulary item format in terms of subtest reliability (0.947 versus 0.927), criterion-related validity (0.820 versus 0.793), and mean item discriminability (0.414 versus 0.378); however, the differences were not found to be statistically significant for the present sample of persons and items. Moreover, use of a reduced-length stem with multiple-choice matching format that uniformly provided inference-generating information in the sentence stem tended to surpass current TOEFL vocabulary item format in terms of criterion-related validity (0.806 versus 0.793); however, this difference was not statistically significant.

The 75-minute familiarization activity employed in the present study did not appear to affect subsequent performance differentially on any of the item format types investigated. This lack of effect may have been due to the nature of the duration, intensity, or focus of the familiarization activity. It may also be the case, and appears more likely, that all of the item formats investigated are similarly impervious to familiarization effects for students similar to those in the present sample. The design of future studies involving familiarization activities could permit an increase in the exposure time in order to ascertain whether more extensive familiarization treatments would produce some effect. Self-reports of familiarity with the respective item format types did indicate that subjects were more familiar with some item formats than with others (Table 4). The current TOEFL vocabulary item format was reportedly the second most familiar

format for these subjects. The three most familiar item formats exhibited a significant positive correlation between self-reported familiarity and performance with those item formats (Table 5). Such correlations, however, do not lend themselves well to causal interpretations. It is likely that more proficient students had more exposure to English language instruction and assessment, with the possible result that they were more likely to have been exposed to the most common item formats. To argue that their differentially greater exposure to those three formats caused them to perform better with those formats would seem to be less defensible, even though the most familiar formats were also found to be the easiest formats.

There was a significant tendency for matching-type items to be easier than comparable supply-type items. While matching-type items also tended to surpass comparable supply-type multiple-choice items in terms of subtest reliability, criterion-related validity, and mean item discriminability, these differences were not uniformly significant statistically. Use of single word or phrase multiple-choice matching format tended consistently by all considered criteria to be inferior to multiple-choice matching formats that employed complete sentence contextualization of target vocabulary in the item stem. The contribution of context both to the clarification and to the inferencing of meaning may underlie the superiority of contextualized formats. Use of such multiple-choice contextualized formats may, for similar reasons, be preferable to use of noncontextualized formats, even when the noncontextualized formats do not involve multiple-choice type tasks (Meara & Buxton, 1987).

These results suggest that future revisions of the TOEFL vocabulary section that entail the embedding of item stems in reading passages in the manner done in the present study, or that incorporate inference-generating information in the sentence stem, would not impair certain estimations of test reliability or validity, and may even enhance them. This conclusion regarding use of extended context for vocabulary testing also gathers support from the findings of Schedl and Way (1990). The presumed advantage of matching over supply-type vocabulary item format may relate to a tendency of supply-type item format to call upon knowledge of syntax, in addition to the semantic knowledge usually required in a vocabulary test. There is, however, abiding controversy concerning which factors in the reading process outweigh others in producing accurate or faulty meaning constructions (Alderson, 1984; Perkins & Linnville, 1987; Sarig, 1989). Such controversy should be resolved through appropriate research studies designed for that purpose.

References

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. Urquhart (Eds.), Reading in a foreign language. London: Longman.

De Jong, H. A. L., and Henning, G. (1990). Test dimensionality in relation to student proficiency. A paper presentation at the Twelfth Annual Language Testing Research Colloquium, San Francisco, March 3-5.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W., and Liskin-Gasparro, J. E. (1985). TOEFL from a communicative viewpoint on language proficiency: A working paper (TOEFL Research Report 17). Princeton, NJ: Educational Testing Service.

Einot, I., and Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. Journal of the American Statistical Association, 70, 574-583.

Gulliksen, H. (1987). Theory of mental tests. Hillsdale, NJ: Lawrence Erlbaum Associates.

Meara, P., and Buxton, B. (1987). An alternative to multiple-choice vocabulary tests. Language Testing, 4(2), 142-154.

Oller, J. W., Jr. (1979). Language tests at school. London: Longman.

Perkins, K., and Linnville, S. E. (1987). A construct definition study of a standardized ESL vocabulary test. Language Testing, 4(2), 125-141.

Pike, L. W. (1979). An evaluation of alternative item formats for testing English as a foreign language (TOEFL Research Report 2). Princeton, NJ: Educational Testing Service.

Sarig, G. (1989). Testing meaning construction: Can we do it fairly? Language Testing, 6(1), 77-94.

Schedl, M., and Way, W. D. (1990). TOEFL vocabulary items: The impact of context. A paper presented at the ETS DIF Interest Group seminar, June 12. Princeton, NJ: Educational Testing Service.

Stansfield, C. W. (1986). Toward communicative competence testing: Proceedings of the second TOEFL invitational conference (TOEFL Research Report 21). Princeton, NJ: Educational Testing Service.

Valette, R. M. (1977). Modern language testing. (2nd ed.) New York: Harcourt, Brace and Jovanovich.

Williams, R., and Dallas, D. (1984). Aspects of vocabulary in the readability of content area L2 educational textbooks: A case study. In J. C. Alderson and A. H. Urquhart (Eds.), Reading in a foreign language. London: Longman.

Wilson, K. M. (1982). A comparative analysis of TOEFL examinee characteristics, 1977-1979 (TOEFL Research Report 11). Princeton, NJ: Educational Testing Service.

Table 1

Sample Description (N = 179)*

Native Language		Sex		Age in Years		Length of Stay in English-Speaking Country	
Arabic	2	Female	72	15-20 =	35	0-6 months =	132
Catalan	1	Male	107	21-25 =	86	6-12 months =	34
Chinese	15			26-30 =	41	13-18 months =	5
Czech	1			31-35 =	11	19-24 months =	0
French	4			36-40 =	1	2-3 years =	4
German	3			41-45 =	2	4-5 years =	1
Hebrew	1			46-50 =	3	Not reported =	3
Indonesian	3						
Italian	4						
Japanese	47						
Korean	60						
Persian	2						
Russian	1						
Spanish	23						
Thai	7						
Turkish	5						
TOTAL	179		179		179		179

*An additional 11 subjects completed the familiarization task but did not participate in experimental testing.

Table 2

Experimental Design

Day	N	Tasks	Minutes	Items Per Person	Item Types	Items Per Item Type	Test Forms
1	99	Instructions	10				
		Feedback	5				
		Familiarization	60	80	8	10	1
2	179	Instructions	10				
		Demographic Questionnaire	5	4			
		Familiarity Questionnaire	10	8			
		Vocabulary Test	90	120	8	15	8

Person Group	N	Item Group Assignments (Day 2)							
1	24	A1	B2	C3	D4	E5	F6	G7	H8
2	24	B1	C2	D3	E4	F5	G6	H7	A8
3	24	C1	D2	E3	F4	G5	H6	A7	B8
4	24	D1	E2	F3	G4	H5	A6	B7	C8
5	21	E1	F2	G3	H4	A5	B6	C7	D8
6	22	F1	G2	H3	A4	B5	C6	D7	E8
7	19	G1	H2	A3	B4	C5	D6	E7	F8
8	21	H1	A2	B3	C4	D5	E6	F7	G8

*Person groups ranged in size from 19 to 24 and were determined by spiraled dissemination of tests at each site. Each group received a different test form.

*Item group letters (A-H) correspond to clusters of 15 distinct vocabulary items and their associated distractors.

*Item group numbers (1-8) correspond to item format types.

*Total items equaled 1,040 (960 experimental items plus 80 familiarization items).

Table 3

Means, Standard Deviations, and Internal Consistency Reliabilities of All Tests (N = 179 Persons)

Item Type	N Items	Mean*	SD*	KR21**
1. Current TOEFL Format	15(8) = 120	8.335	2.941	.927
2. Word or Phrase, Matching	15(8) = 120	7.749	2.677	.893
3. Short Sentence, Matching	15(8) = 120	8.821	2.721	.906
4. Short Sentence, Supply	15(8) = 120	7.648	2.507	.857
5. Inference Sentence, Matching	15(8) = 120	8.430	2.804	.913
6. Inference Sentence, Supply	15(8) = 120	6.832	2.706	.899
7. Word or Phrase, Embedded Options	15(8) = 120	7.380	2.768	.906
8. Passage-Embedded Sentence, Matching	15(8) = 120	7.419	3.244	.947

*Mean of eight clusters of 15 items per cluster.

**120-item Spearman-Brown projections from the means of eight 15-item clusters.

Table 4

Familiarity Means, Standard Deviations, and Correlations with Scores on Eight Item Types (N = 179)

Item Type	Mean	S.D.	R
1. Current TOEFL Format	1.936	1.015	.242**
2. Word or Phrase, Matching	1.407	0.996	-.057
3. Short Sentence, Matching	1.843	0.964	.182*
4. Short Sentence, Supply	1.738	0.953	-.170
5. Inference Sentence, Matching	2.023	0.911	.219*
6. Inference Sentence, Supply	1.105	1.021	-.047
7. Word or Phrase, Embedded Options	1.157	1.028	-.068
8. Passage-Embedded Sentence, Matching	0.866	0.967	-.078
9. Total of All Item Types	12.076	4.604	.093

*p < .05
**p < .01

Means and standard deviations are based on questionnaire self-report of familiarity on a 0-3-point scale.

R = Pearson product-moment correlation of reported familiarity, with scores on associated item types.

Table 5

Pearson Product-Moment Correlations Among All Item Types (AT through HT), Total Vocabulary Scores (T), Familiarization Activity (FA), and Self-Reported Familiarity within Individual Item Types (Q)
(N = 172)

	AT	BT	CT	DT	ET	FT	GT	HT	T	FA
AT	1.000									
BT	0.511	1.000								
CT	0.650	0.470	1.000							
DT	0.506	0.372	0.489	1.000						
ET	0.665	0.518	0.603	0.496	1.000					
FT	0.340	0.397	0.464	0.392	0.421	1.000				
GT	0.513	0.445	0.488	0.471	0.568	0.508	1.000			
HT	0.575	0.503	0.561	0.581	0.583	0.465	0.616	1.000		
T	0.793	0.695	0.782	0.708	0.806	0.658	0.765	0.820	1.000	
FA	-0.083	-0.109	0.015	-0.032	-0.085	-0.099	-0.025	-0.025	-0.075	1.000
Q	0.242	-0.057	0.182	-0.170	0.219	-0.047	-0.068	-0.078	0.093	0.151

AT = Current TOEFL vocabulary item format
BT = Isolated word or phrase, multiple-choice matching format
CT = Minimal sentence stem, multiple-choice matching format
DT = Minimal sentence stem, multiple-choice cloze-supply format
ET = Reduced-length inference-generating sentence stem, multiple-choice matching format
FT = Reduced-length inference-generating sentence stem, multiple-choice cloze-supply format
GT = Single word or phrase stimulus with passage-embedded multiple-choice options
HT = Passage-embedded sentence stem, multiple-choice matching format

Table 6A

ANOVA with Item Difficulty Delta as the Dependent Variable
and Item Format as the Independent Variable
(N = 960 items)

Source	DF	Mean Square	F	P
Constant	1			
Item Format	7	81.139	5.244	<0.001
Error	952	15.473		
Total	960			

Multiple R = .192

Table 6B

ANOVA with Item Discriminability Fisher Z Transformed Point Biserial r as
the Dependent Variable and item Format as the Independent Variable
(N = 960 items)

Source	DF	Mean Square	F	P
Constant	1			
Item Format	7	0.322	4.375	<0.001
Error	952	0.074		
Total	960			

Multiple R = .176

Table 7

Item Difficulty and Item Discriminability Statistics (Delta and Point Biserial Z) for 960 Vocabulary Items by Item Format Type

Item Type	N Items	Item Difficulty		Item Discriminability	
		Delta	S.D.	Mean Point Biserial Z	SD
1. Current TOEFL Format	120	12.304	4.108	.378	.222
2. Word or Phrase, Matching	120	13.094	4.160	.298	.316
3. Short Sentence, Matching	120	11.779	4.164	.367	.282
4. Short Sentence, Supply	120	13.159	3.759	.271	.258
5. Inference Sentence, Matching	120	12.268	3.820	.372	.289
6. Inference Sentence, Supply	120	14.216	3.636	.280	.246
7. Word or Phrase, Embedded Options	120	13.651	3.965	.336	.294
8. Passage-Embedded Sentence, Matching	120	13.534	3.819	.414	.250

Tukey's HSD 0.05 critical range for pairs of Delta means = 1.539
Tukey's HSD 0.05 critical range for pairs of point biserial means = .106

Appendix A: Sample Test and Answer Sheet

FORM A1

Directions: Each sentence has an underlined word or phrase. Below each sentence there are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

1. It has been observed that people generally make the best of a bad situation when they know that it is sure to occur.

   (A) likely
   (B) soon
   (C) certain
   (D) advisable

2. The harp is one of the most ancient types of instrument still in use.

   (A) earliest
   (B) strangest
   (C) most expensive
   (D) most intricate

3. Orchids have an impressive ecological distribution when the whole plant family is considered.

   (A) neighboring
   (B) wild
   (C) huge
   (D) entire

4. The task of teaching is central to education.

   (A) job
   (B) amount
   (C) joy
   (D) philosophy

5. Almost any process used for molding metals may also be used to shape plastics.

   (A) Nearly
   (B) Presumably
   (C) Fortunately
   (D) Predictably

6. Combinations of tones sound pleasant to the ear when the individual frequencies are in simple proportion.

(A) loud
(B) agreeable
(C) pompous
(D) static

7. Dryness in the nose and throat is a side effect often associated with antihistamines.

(A) major symptom
(B) secondary reaction
(C) rapid cure
(D) desired response

8. During his lifetime, Herman Melville's celebrated novel "Moby Dick" was not recognized as great literature.

(A) lengthy
(B) exotic
(C) famous
(D) maritime

9. With the aid of new technology and innovative procedures, meteorologists have enhanced their understanding of atmospheric conditions and weather systems.

(A) improved
(B) analyzed
(C) tested
(D) debated

10. Since crystals have characteristic surfaces and shapes, crystallography can be a valuable tool in mineral identification.

(A) angular
(B) complex
(C) infinite
(D) distinctive

11. The corporate gymnasium provides employees with a healthy outlet for tension and stress.

(A) discussion on
(B) approach to
(C) schedule for
(D) release from

12. In her late seventies, too frail to continue her arduous farm work, Anna Mary Moses began painting simple scenes of rural life.

(A) busy
(B) poor
(C) bored
(D) weak

13. Vivid garments with exotic motifs attract the attention of tourists in Hawaii.

(A) posters
(B) dances
(C) restaurants
(D) clothes

14. Barnacles use secretions to cement their shells to underwater objects.

(A) bring
(B) point
(C) bond
(D) contrast

15. Although many national leaders objected strenuously to the acquisition of Alaska, Secretary of State William Seward managed to secure its purchase in 1867.

(A) privately
(B) initially
(C) justifiably
(D) vigorously

FORM B2

Directions: Choose the word (marked A, B, C or D) that best keeps the meaning of the underlined word or phrase. Mark your choice on your answer sheet.

****

16. elected
    (A) paid
    (B) advised
    (C) consulted
    (D) chosen

17. thrust
    (A) collision
    (B) angle
    (C) push
    (D) erosion

18. jettison
    (A) discard
    (B) sell
    (C) destroy
    (D) add

19. rather than
    (A) restricted
    (B) as well as
    (C) unless
    (D) instead of

20. most attractive
    (A) prettiest
    (B) rarest
    (C) softest
    (D) strongest

21. cloth
    (A) fabric
    (B) clothing
    (C) tapestry
    (D) thread

22. quest
    (A) redundancy in
    (B) conflict of
    (C) quality in
    (D) pursuit of

23. an academy
    (A) a philosophy
    (B) a clinic
    (C) a school
    (D) a company

24. creek
    (A) trench
    (B) gorge
    (C) ravine
    (D) stream

25. suppress
    (A) check
    (B) cause
    (C) demonstrate
    (D) worsen

26. arouse
    (A) excite
    (B) harass
    (C) appease
    (D) arrest

27. now
    (A) only
    (B) at present
    (C) barely
    (D) without question

28. officially
    (A) arbitrarily
    (B) formally
    (C) frankly
    (D) internally

29. cultivated
    (A) seen
    (B) grown
    (C) dried
    (D) cooked

30. sharply
    (A) grimly
    (B) steeply
    (C) constantly
    (D) roughly

FORM C3

Directions: Each sentence has an underlined word or phrase. Below each sentence are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

31. People prefer San Diego's mild climate.

(A) biological balance
(B) rotation
(C) weather conditions
(D) geography

32. Hungry goats feed on anything.

(A) hunt
(B) eat
(C) keep
(D) see

33. In fact, he is my brother.

(A) credibly
(B) actually
(C) reportedly
(D) potentially

34. The fisherman decided to take advantage of the good weather.

(A) reinstall
(B) make use of
(C) augment
(D) make sense of

35. Samuel found the answer singlehandedly.

(A) by himself
(B) at one time
(C) manually
(D) boldly

36. The dim light was not adequate.

(A) weak
(B) static
(C) harsh
(D) deep

37. His lack of training caused many blunders.

(A) mistakes
(B) attempts
(C) insults
(D) arguments

38. Illiteracy causes barriers to learning.

(A) amendments to
(B) regulations of
(C) obstacles to
(D) divisions of

39. The disparaging remarks hurt him.

(A) authoritative
(B) lengthy
(C) uninteresting
(D) negative

40. Satellites may be propelled by solar wind.

(A) carved
(B) docked
(C) driven forward
(D) carried upright

41. The marathon was grueling.

(A) exhilarating
(B) gruesome
(C) exhausting
(D) creative

42. He erupted in <u>an outburst</u> of criticism.

(A) an epidemic
(B) an explosion
(C) a range
(D) a disruption

43. Cigarettes are <u>harmful</u>.

(A) detrimental
(B) prodigious
(C) intrusive
(D) mordant

44. Some cars are <u>inherently</u> bad.

(A) relatively
(B) intrinsically
(C) sporadically
(D) incrementally

45. It is the same <u>throughout</u> the world.

(A) elsewhere
(B) down under
(C) all over
(D) out of

FORM D4

Directions: Choose the best word to complete the most common meaning of each sentence.

****

46. They emigrated to this _____ from abroad.

(A) nation
(B) county
(C) terrain
(D) continent

47. Einstein was a _____ scientist.

(A) kindly
(B) noble
(C) famous
(D) dashing

48. It brought rapid _____ in the economy.

(A) furor
(B) nurturing
(C) expansion
(D) patenting

49. He _____ ever leaves his room.

(A) sadly
(B) rarely
(C) quietly
(D) briefly

50. The workers _____ grim houses.

(A) longed for
(B) resided in
(C) planned for
(D) worshipped in

51. The army tried to _____ the training.

(A) change
(B) speed up
(C) turn around
(D) destroy

52. The secret code was not _____ to others.

(A) acceptable
(B) understandable
(C) interesting
(D) pleasing

53. In democracy rule is by the _____ of the people.

(A) majority
(B) assignments
(C) texts
(D) rest

54. The engineer _____ an ingenious invention.

(A) demanded
(B) created
(C) diagnosed
(D) allowed

55. The wind caused the sail to be _____.

(A) lightly tinted
(B) somewhat opaque
(C) tightly stretched
(D) delicately made

56. The earthquake was caused by a _____ in the crust.

(A) a cave
(B) an indentation
(C) a crack
(D) a hole

57. They left together to travel _____.

(A) smoothly
(B) concurrently
(C) effectively
(D) spontaneously

58. She recognized all the _____ details.

(A) short-lived
(B) dried-up
(C) minuscule
(D) minor

59. Her talent served to _____ the audience.

(A) injure
(B) infuriate
(C) daze
(D) exhaust

60. He hoped to _____ the beautiful girl.

(A) intrigue
(B) court
(C) think little of
(D) try to avoid

FORM E5

Directions: Each sentence has an underlined word or phrase. Below each sentence there are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

61. Although starting the railroad was simple, completing it was not simple.

(A) inspecting
(B) suggesting
(C) complementing
(D) finishing

62. Study in the field of transportation was most interesting to her.

(A) area
(B) cargo
(C) lecture about
(D) argue about

63. The harsh chemical did not treat the metal gently.

(A) continuously
(B) comically
(C) routinely
(D) softly

64. The shy scientist could not deal with the worldwide recognition.

(A) unexpected
(B) immediate
(C) international
(D) temporary

65. The noisy dormitory made it difficult for the students to concentrate on their studies.

(A) count on
(B) focus on
(C) lecture about
(D) argue about

66. Although money was scarce, they had an endless supply of bills to pay.

(A) measurable
(B) limitless
(C) worthless
(D) infinitesimal

67. Before he took off, the pilot was absolutely sure the plane was safe.

(A) totally
(B) often
(C) intrinsically
(D) irresistibly

68. He fascinated his teachers with his astounding memory.

(A) alarming
(B) annoying
(C) amazing
(D) alluring

69. An independent person, he objected when the army made him forsake his identity.

(A) enter
(B) make use of
(C) go to the edge of
(D) leave

70. The former President was a prominent figure at the convention.

(A) an insightful
(B) an interesting
(C) a determined
(D) a notable

71. The Swiss are known for their quality timepieces.

(A) works of art
(B) spare parts
(C) calendars
(D) clocks

72. A folk song will often vary its rhythm or alternate between major and minor keys.

(A) lyrics
(B) notes
(C) beat
(D) message

73. Since he developed the car, he was keenly aware of its shortcomings.
(A) sadly
(B) currently
(C) anxiously
(D) acutely

74. The red bow joined the baby's wispy strands of hair.

(A) crispy
(B) delicate
(C) lofty
(D) transparent

75. Despite the gardener's threats, the two cows roamed unrestrained toward his garden.

(A) in masses
(B) aimlessly
(C) enthusiastically
(D) at will

FORM F6

Directions: Choose the best word to complete each sentence. If more than one word is possible, choose the word or expression that gives the most common meaning.

****

76. He may have won the _____ but he lost the war.

(A) flight
(B) struggle
(C) option
(D) guess

77. While Carter is on that team, Brooks and Samuel _____ another team.

(A) are bigger than
(B) are contrasted with
(C) are members of
(D) are nourishment for

78. After the picnic, the park was littered with _____.

(A) filler
(B) garbage
(C) paper
(D) fuel

79. Joe admitted the game was lost but Al refused to _____ the fact.

(A) confront
(B) mock
(C) lose
(D) appreciate

80. Rather than dispose of the cans, he gathered them for _____.

(A) cleaning
(B) reprocessing
(C) reselling
(D) crushing

81. Although they failed to get the contract, the owner appreciated their _____.

(A) requests
(B) attempts
(C) offers
(D) trips

82. The rich harvest provided _____ amount of food.

(A) a solid
(B) an indefinite
(C) a plentiful
(D) an elaborate

83. After the storm, the river _____ the town.

(A) flooded
(B) eroded
(C) fertilized
(D) seeded

84. Because he knew the area well, he was _____ more help to the searchers than strangers were.

(A) much
(B) even
(C) no
(D) hardly

85. Jennings may avoid paying taxes for a while, but _____ the government will catch him.

(A) in the final analysis
(B) in a predictable way
(C) unthinkingly
(D) religiously

86. Although the Congress approved the bill, the President threatened to be the main _____ its passage.

(A) factor in
(B) impediment to
(C) occurrence in
(D) phenomenon of

87. The senator decided to vacation for a few days so that afterwards he could energetically _____ reelection.

(A) decide against
(B) campaign for
(C) anticipate
(D) consider

88. Although they were twins, they had _____ personalities.

(A) split
(B) equal
(C) detachable
(D) distinct

89. Make no mistake; Jack does appear inept, but he's really _____ clever.

(A) misleadingly
(B) generally
(C) noticeably
(D) exceptionally

90. Paying taxes is _____ event for most citizens.

(A) an average
(B) a nearby
(C) a centrally located
(D) an unusually bright

FORM G7

Directions: Find the underlined word, A, B, C, or D, in each sentence that is closest in meaning to the word that is given.

****

91. included

(A) (B)
Once the agent deciphered the message discovered on the photograph, he
(C) (D)
questioned the spy to get the meaning of what the note contained.

92. particular

(A) (B)
The seasoned veteran ball player was biased against all rookies, but
(C) (D)
especially against a certain shortstop with fastidious fielding.

93. handsome

(A) (B) (C)
The well-worn cabinet appeared naked until the carpenter completed the
(D)
application of the good-looking finish.

94. draw

(A) (B)
Johnson felt he would delight the owners if he could inspire his players
(C)
to do better and attract larger crowds, but even this effort could not
(D)
console the disenchanted management.

95. endeavoring

(A) (B)
While Ed was trying to prepare himself for requesting a raise, the boss
(C) (D)
was deciding against offering anyone more money.

96. actually

(A) (B) (C)
Supposedly, we still have salespersons who can really succeed
(D)
everywhere.

97. consciously

(A)
Conveniently, the worker feigned sickness to justify his being away
(B) (C) (D)
intentionally, so now we should reprimand him robustly and lastingly
curtail such behavior.

98. acclaim

(A) (B)
With confidence and good training, the accountant requested
(C) (D)
reimbursement for his expenses in place of recognition.

99. placid

(A)
As reporters gathered on the hilly mountainside, the expert remained
(B) (C)
calm and announced that the painting discovered in the seedy cabin was
(D)
undeniably a fake.

100. restrict

(A) (B) (C)
The judge refused to limit the efforts to standardize the test, and he
(D)
agreed to prohibit its publication.

101. curative

(A) (B) (C)
The superior healing ability of these organic medicines seems almost
(D)
magical.

102. bore

(A) (B)
He decided to slice a section of plastic to trim the doorway but when
(C) (D)
he began to drill, the material started to melt.

103. take on

(A) (B)
The fugitive will assume a false identity in order to avoid detection,
(C) (D)
and he will regulate his driving so as not to exceed the speed limit.

104. a curiously

(A) (B)
An overly zealous manager and an unfortunately inexperienced clerk
(C) (D)
transformed an unambiguously worded order into an oddly confusing
document.

105. mainly

(A) (B) (C)
Likewise, that somewhat strange response was merely a first reaction and
(D)
was mostly ignored by the staff.

FORM H8

Directions: Read the following passage. For each of the numbered and underlined words or phrases in the passage, find the word, A, B, C, or D, in each group below that is closest in meaning to the word found in the passage.

****

Neil Armstrong (106) calmly grasped the controls of the lunar module and began to (107) scan the white surface below him. He had (108) hardly looked down through the windows when he noticed huge rocks strewn across the landing area (109) selected by remote computers. Bathed in the (110) luminous aura of the craft's instruments, Armstrong (111) cut the module's rate of descent and (112) retreated from the original landing site. The (113) brave astronaut continued to traverse the area as his fuel fell below the (114) optimal level. His ship moved (115) erratically until it finally put down on a clear area.

The (116) duration of the stay was (117) slightly more than a few hours for collecting soil samples, conducting experiments, and photographing the stark (118) grandeur of the lunar surface. After these (119) arduous duties were completed, he rejoined the mother ship. But all of this might never have occurred were it not for some creative piloting (120) prior to the first human step on the moon.

106. (A) considerately
(B) hungrily
(C) alertly
(D) quietly

107. (A) survey
(B) strip
(C) smell
(D) chart

108. (A) carefully
(B) probably
(C) scarcely
(D) slowly

109. (A) rumored
(B) supposed
(C) chosen
(D) willing

110. (A) ludicrous
(B) glowing
(C) magnetic
(D) flickering

111. (A) conceal
(B) review
(C) damage
(D) reduce

112. (A) received praise from
(B) withdrew from
(C) rebelled against
(D) traveled around

113. (A) thin
(B) hearty
(C) curious
(D) courageous

114. (A) most advantageous
(B) most convenient
(C) most flexible
(D) most obvious

115. (A) endlessly
(B) irregularly
(C) ominously
(D) longitudinally

116. (A) length
(B) severity
(C) frequency
(D) penetration

117. (A) no
(B) a little
(C) even
(D) very much

118. (A) delicacy
(B) turbulence
(C) scenery
(D) magnificence

119. (A) rewarding
(B) difficult
(C) vital
(D) final

120. (A) before
(B) in
(C) at the time of
(D) due to

Appendix B: Familiarization Materials

FORM Q1

Directions: Each sentence has an underlined word or phrase. Below each sentence there are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

1. The nuthatch gets food by wedging a nut into a crevice in a tree trunk and then opening the nut with its sharp beak.

   (A) cramming
   (B) splicing
   (C) wending
   (D) beating

2. Platinum is neither corroded by moisture nor affected by oxygen or ordinary acids.

   (A) polished
   (B) softened
   (C) taken over
   (D) worn away

3. The coming of the railroads to Minnesota was accompanied by a surge in the rate of industrial expansion in that state.

   (A) lull
   (B) shift
   (C) jump
   (D) cut

4. Volcanic pumice originates in the frothy part of lava.

   (A) greasy
   (B) foamy
   (C) fluid
   (D) toxic

5. Treasurer of the United States Katherine Ortega delivered the keynote address at the 1984 Republican National Convention in Dallas, Texas.

   (A) bid
   (B) speech
   (C) envelope
   (D) nomination

6. Safety officials and industry representatives have earnestly questioned whether the increased use of synthetic materials heightens the risk of fire.

(A) seriously
(B) cautiously
(C) secretly
(D) continually

7. Artist Isamu Noguchi is best known for his abstract sculptures designed as adjuncts to architectural works.

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

8. Eating foods that are high in cholesterol and fats may affect the health of an individual deleteriously.

(A) superficially
(B) dramatically
(C) adversely
(D) latently

9. After five years of dogged effort, Charles Goodyear invented the process of vulcanization, which made the widespread use of rubber possible.

(A) vague
(B) inspiring
(C) fruitless
(D) persistent

10. Herbal medicine has been used through the ages for treating ailments.

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q2

Directions: Choose the word (marked A, B, C or D) that best keeps the meaning of the underlined word or phrase. Mark your choice on your answer sheet.

****

11. wedging

(A) cramming
(B) splicing
(C) wending
(D) beating

12. corroded

(A) polished
(B) softened
(C) taken over
(D) worn away

13. surge

(A) lull
(B) shift
(C) jump
(D) cut

14. frothy

(A) greasy
(B) foamy
(C) fluid
(D) toxic

15. address

(A) bid
(B) speech
(C) envelope
(D) nomination

16. earnestly

(A) seriously
(B) cautiously
(C) secretly
(D) continually

17. adjuncts to

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

18. deleteriously

(A) superficially
(B) dramatically
(C) adversely
(D) latently

19. dogged

(A) vague
(B) inspiring
(C) fruitless
(D) persistent

20. ailments

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q3

Directions: Each sentence has an underlined word or phrase. Below each sentence there are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

21. He was wedging food into his mouth.

(A) cramming
(B) splicing
(C) wending
(D) beating

22. The old iron pan was corroded by moisture.

(A) polished
(B) softened
(C) taken over
(D) worn away

23. Adrenalin gave the athletes a surge in energy.

(A) lull
(B) shift
(C) jump
(D) cut

24. The white surface of the soapy water appeared frothy.

(A) greasy
(B) foamy
(C) fluid
(D) toxic

25. She gave the keynote address.

(A) bid
(B) speech
(C) envelope
(D) nomination

26. He earnestly warned the child of the falling object.

(A) seriously
(B) cautiously
(C) secretly
(D) continually

27. They made adjuncts to their garden space to increase production.

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

28. Most toxins affect health deleteriously.

(A) superficially
(B) dramatically
(C) adversely
(D) latently

29. They finally succeeded by dogged effort.

(A) vague
(B) inspiring
(C) fruitless
(D) persistent

30. Medical science attempts to eliminate ailments.

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q4

Directions: Choose the best word to complete the most common meaning of each sentence.

****

31. He was _____ food into his mouth.

(A) cramming
(B) splicing
(C) wending
(D) beating

32. The old iron pan was _____ by moisture.

(A) polished
(B) softened
(C) taken over
(D) worn away

33. Adrenalin gave the athletes a _____ in energy.

(A) lull
(B) shift
(C) jump
(D) cut

34. The white surface of the soapy water appeared _____.

(A) greasy
(B) foamy
(C) fluid
(D) toxic

35. She gave the keynote _____.

(A) bid
(B) speech
(C) envelope
(D) nomination

36. He _____ warned the child of the falling object.

(A) seriously
(B) cautiously
(C) secretly
(D) continually

37. They made _____ their garden space to increase production.

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

38. Most toxins affect health _____.

(A) superficially
(B) dramatically
(C) adversely
(D) latently

39. They finally succeeded by _____ effort.

(A) vague
(B) inspiring
(C) fruitless
(D) persistent

40. Medical science attempts to eliminate _____.

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q5

Directions: Each sentence has an underlined word or phrase. Below each sentence there are four other words or phrases, marked (A), (B), (C), and (D). You are to choose the one word or phrase that best keeps the meaning of the original sentence if it is substituted for the underlined word or phrase. Then, on your answer sheet, find the number of the question and fill in the space that corresponds to the letter you have chosen. Fill in the space so that the letter inside the oval cannot be seen.

****

41. By wedging the envelope into the crack beneath the door he was able to pass the message outside the room.

    (A) cramming
    (B) splicing
    (C) wending
    (D) beating

42. Time and moisture had completely corroded the old iron spear.

    (A) polished
    (B) softened
    (C) taken over
    (D) worn away

43. The soldiers' rising adrenalin level gave them a surge in energy that enabled them to meet the challenge.

    (A) lull
    (B) shift
    (C) jump
    (D) cut

44. Agitating the soapy water produced a frothy layer on the surface.

    (A) greasy
    (B) foamy
    (C) fluid
    (D) toxic

45. The audience applauded after she gave the keynote address.

    (A) bid
    (B) speech
    (C) envelope
    (D) nomination

46. The blind man's friends earnestly warned him about the object that was falling from the sky.

(A) seriously
(B) cautiously
(C) secretly
(D) continually

47. Using all available external space, they made adjuncts to their office space to permit growth.

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

48. The food she ate affected her health deleteriously, so she eventually had to see a physician.

(A) superficially
(B) dramatically
(C) adversely
(D) latently

49. Dogged effort will often succeed after much time even when other approaches fail.

(A) Vague
(B) Inspiring
(C) Fruitless
(D) Persistent

50. By appropriate treatment it is possible to eliminate most of the ailments known to humanity.

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q6

Directions: Choose the best word to complete each sentence. If more than one word is possible, choose the word or expression that gives the most common meaning.

****

51. By _____ the envelope into the crack beneath the door he was able to pass the message outside the room.

(A) cramming
(B) splicing
(C) wending
(D) beating

52. Time and moisture had completely _____ the old iron spear.

(A) polished
(B) softened
(C) taken over
(D) worn away

53. The soldiers' rising adrenalin level gave them a _____ in energy that enabled them to meet the challenge.

(A) lull
(B) shift
(C) jump
(D) cut

54. Agitating the soapy water produced a _____ layer on the surface.

(A) greasy
(B) foamy
(C) fluid
(D) toxic

55. The audience applauded after she gave the keynote _____.

(A) bid
(B) speech
(C) envelope
(D) nomination

56. The blind man's friends _____ warned him about the object that was falling from the sky.

(A) seriously
(B) cautiously
(C) secretly
(D) continually

57. Using all available external space, they made _____ their office space to permit growth.

(A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

58. The food she ate affected her health _____, so she eventually had to see a physician.

(A) superficially
(B) dramatically
(C) adversely
(D) latently

59. _____ effort will often succeed after much time even when other approaches fail.

(A) Vague
(B) Inspiring
(C) Fruitless
(D) Persistent

60. By appropriate treatment it is possible to eliminate most of the _____ known to humanity.

(A) doctors
(B) diseases
(C) patients
(D) livestock

FORM Q7

Directions: Find the underlined word, A, B, C, or D, in each sentence that is closest in meaning to the word that is given.

****

61. wedging

Here the work of a telephone worker consisted mainly of (A) wending one's way through the jungle, (B) beating back wild overgrowth, (C) cramming sharpened poles into the soft soil, and (D) splicing electrical wires together.

62. corroded

They (A) polished the old anchor as best they could, but found that is was largely (B) taken over by layers of (C) softened barnacles and (D) worn away by the elements.

63. surge

There was no (A) lull in the storm but only an occasional (B) shift in direction and a sudden (C) jump in force as the tornado (D) cut across the valley.

64. frothy

They used a (A) foamy detergent and (B) fluid cleaners to remove the (C) toxic residue from the (D) greasy surfaces of the vats.

65. address

The (A) envelope contained the text of the (B) speech she would give in her (C) bid for the (D) nomination.

66. earnestly

(A) (B)
Secretly they acknowledged that they continually needed to proceed
(C) (D)
cautiously if they seriously hoped to succeed.

67. adjuncts to

(A) (B)
Advertisements of their office complex claimed advancements in
(C)
architectural design, but actually they had made only a few adaptations of
(D)
the previous plan by creating additions to an older building.

68. deleteriously

(A) (B)
Some medications affect health only superficially or even adversely, while
(C)
other treatments reduce symptoms dramatically or else produce results
(D)
latently.

69. dogged

(A) (B)
He awoke with a vague, inspiring feeling that unless he fought on with
(C) (D)
persistent effort, all his hopes would be fruitless.

70. ailments

(A) (B) (C) (D)
Doctors combat diseases in both human patients and in livestock.

FORM Q8

Directions: Read the following passage. For each of the numbered and underlined words or phrases in the passage, find the word, A, B, C, or D, in each group below that is closest in meaning to the word found in the passage.

****

We sometimes take for granted the contributions of science and technology in reducing physical ailments[71] or in providing conveniences like the automobile and the airplane. We tend to forget the technicians working earnestly[72] with dogged[73] determination under conditions that may affect their own health deleteriously[74] to provide us with these advantages. Whether chemists working with frothy[75] chemicals in the isolated adjuncts to[76] their laboratories, or aeronautical engineers wedging[77] strips of some not easily corroded[78] alloy into the frame of a weather satellite, all have contributed to the surge[79] in scientific knowledge. We may never attend a meeting of a scientific society to hear some address[80] on the latest breakthroughs, but we have all benefited from scientific endeavor.

71. (A) doctors
(B) diseases
(C) patients
(D) livestock

72. (A) seriously
(B) cautiously
(C) secretly
(D) continually

73. (A) vague
(B) inspiring
(C) fruitless
(D) persistent

74. (A) superficially
(B) dramatically
(C) adversely
(D) latently

75. (A) greasy
(B) foamy
(C) fluid
(D) toxic

76. (A) additions to
(B) adaptations of
(C) advertisements of
(D) advancements in

77. (A) cramming
(B) splicing
(C) wending
(D) beating

78. (A) polished
(B) softened
(C) taken over
(D) worn away

79. (A) lull
(B) shift
(C) jump
(D) cut

80. (A) bid
(B) speech
(C) envelope
(D) nomination

Appendix C: Demographic and Familiarity Questionnaires

QUESTIONNAIRE

Your Name: _____

Your Native Language: _____

Sex: Male _____ Female _____

Age: 15-20 _____ 21-25 _____ 26-30 _____

31-35 _____ 36-40 _____ 41-45 _____

46-50 _____ 51-55 _____ 56-60 _____

Length of stay in an English-speaking country:

0-6 Months _____ 6-12 Months _____

13-18 Months _____ 19-24 Months _____

2-3 Years _____ 4-5 Years _____

6-7 Years _____ 8-9 Years _____

10 or more Years _____

For each of the question types below, show how often you have seen or used this kind of question before by circling the right number (3 = very often; 2 = often; 1 = not often; 0 = never).

SAMPLE QUESTION TYPE	VERY OFTEN			NEVER
I. Human facial expressions differ from those of animals in the degree to which they can be deliberately controlled and modified. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
II. Deliberately (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
III. He saw it deliberately. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
IV. He considered it _____. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
V. He was guilty because he did it deliberately. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
VI. He was guilty because he did it _____. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0
VII. Deliberately (A) (B) They both enjoyed the show absolutely, but he X (C) intentionally returned to see if his feelings (D) would be noticeably different the second time.	3	2	1	0
VIII. She carried out the plan deliberately. 1. (A) discovered X (B) executed (C) lifted (D) abolished 2. (A) both (B) noticeably X (C) intentionally (D) absolutely	3	2	1	0

Cover Printed on Recycled Paper

57606-01197 • Y31M 5 • 275702 • Printed in U.S.A.

BEST COPY AVAILABLE